Complexity and Performance Assessment for Data Fusion Systems

March/April 1998

Martin O. Hofmann and Stephen M. Jameson

Lockheed Martin Advanced Technology Laboratories

1 Federal Street, Camden, NJ 08102

ABSTRACT

Demonstrated, documented performance is a prerequisite before a data fusion system may be deployed. Developers and users must be confident about fusion system performance across the full range of operating conditions and scenarios the system is anticipated to encounter. We report on an approach to multi-sensor data fusion performance characterization which systematically explores system performance and quantifies performance degradation at and beyond the limits of the intended application scenarios. A quantitative characterization of the complexity of test scenarios supports our experimental approach to performance assessment. Scenario complexity characterization directs creation and systematic variation of test scenarios and facilitates efficient exploration of the range of relevant fusion scenarios. Data Fusion performance metrics measure the quality of the track picture produced by the data fusion solution and the correctness of the intermediate constituent processing steps. Track picture quality is measured by the accuracy, precision, consistency, and completeness of the fused track picture. Constituent metrics function as "built-in-test" procedures for critical processing steps and reveal causes for sub-optimal performance. They indicate when the fusion system under test operates on a scenario which approaches the limits of its capabilities. We successfully applied the complexity and performance measures described in this paper to the development and validation of the Rotorcraft Pilot's Associate (RPA) Level 1 Sensor Fusion component.

1. INTRODUCTION

Data Fusion Performance Assessment reveals whether a Data Fusion (DF) solution is appropriate for the intended target environment, sensor suite, and computing platform constraints. Performance assessment determines the range of operating conditions under which a DF solution performs at optimal, near-optimal, and degraded levels, and it provides a rational basis for choosing between competing DF solutions. During fusion system development, rigorous performance assessment assists in selecting and tuning algorithms for repeatable, robust, best-of-breed performance.

The target of our performance assessment is the class of object-level, i.e. Joint Directors of Laboratories (JDL) Level 1, multi-sensor multi-target fusion systems. This class of DF systems accepts single-sensor contacts and tracks from multiple sensors and produces a consolidated track picture, ideally consisting of a single smoothed track for each target. Each output track combines the features contributed by all concurrently reporting sensors. The exact variant of multi-sensor fusion, e.g. centralized, hierarchical with or without fused track feedback, etc., is irrelevant to the performance assessment approach described here.

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (FROM - TO)
01-03-1998	Conference Proceedings	xx-xx-1998 to xx-xx-1998

4. TITLE AND SUBTITLE Complexity and Performance Assessment for Data Fusion Systems Unclassified	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Hofmann, Martin O. ; Jameson, Stephen M. ;	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME AND ADDRESS Lockheed Martin Advanced Technology Laboratories 1 Federal Street Camden, NJ08102	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME AND ADDRESS Director, CECOM RDEC Night Vision and Electronic Sensors Directorate, Security Team 10221 Burbeck Road Ft. Belvoir, VA22060-5806	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
APUBLIC RELEASE
,

13. SUPPLEMENTARY NOTES
See Also ADM201041, 1998 IRIS Proceedings on CD-ROM.

14. ABSTRACT
Demonstrated, documented performance is a prerequisite before a data fusion system may be deployed. Developers and users must be confident about fusion system performance across the full range of operating conditions and scenarios the system is anticipated to encounter. We report on an approach to multi-sensor data fusion performance characterization which systematically explores system performance and quantifies performance degradation at and beyond the limits of the intended application scenarios. A quantitative characterization of the complexity of test scenarios supports our experimental approach to performance assessment. Scenario complexity characterization directs creation and systematic variation of test scenarios and facilitates efficient exploration of the range of relevant fusion scenarios. Data Fusion performance metrics measure the quality of the track picture produced by the data fusion solution and the correctness of the intermediate constituent processing steps. Track picture quality is measured by the accuracy, precision, consistency, and completeness of the fused track picture. Constituent metrics function as ?built-in-test? procedures for critical processing steps and reveal causes for sub-optimal performance. They indicate when the fusion system under test operates on a scenario which approaches the limits of its capabilities. We successfully applied the complexity and performance measures described in this paper to the development and validation of the Rotorcraft Pilot?s Associate (RPA) Level 1 Sensor Fusion component.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19. NAME OF RESPONSIBLE PERSON
			Public Release	16	Fenster, Lynn lfenster@dtic.mil
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER International Area Code Area Code Telephone Number 703767-9007 DSN 427-9007

The Data Fusion performance metrics presented in this paper measure the quality of the track picture produced by the DF solution and the correctness of the intermediate constituent processing steps. Track picture quality is measured by the accuracy, precision, consistency, and completeness of the fused track picture. Constituent processing steps, such as update-to-track association, merit their own metrics. Constituent metrics reveal causes for sub-optimal performance and indicate when the DF solution operates on a scenario which approaches the limits of its capabilities.

Realistic multi-sensor multi-target fusion scenarios are difficult enough to preclude perfect fusion system performance in real time in most cases. Complexity metrics quantify the difficulty presented by a specific scenario and provide a basis for explaining and predicting varying levels of fusion system performance. Complexity metrics allow performance comparisons across diverse scenarios and they direct systematic exploration of the capabilities of a fusion solution.

Complexity metrics estimate performance when direct performance measurement is problematic, e.g. during run-time in actual deployment where ground truth is not available. Most of the performance metrics depend on the knowledge of ground truth, which is available only during simulation runs or from an instrumented test range. Performance metrics calculated without ground truth are less reliable. Complexity assessment compensates for the reduced value of the performance metrics and assists in detecting when fusion performance declines and/or fusion system tuning becomes advisable.

The complexity metrics presented in this paper characterize and quantify the difficulty inherent in the ground truth, i.e. the arrangement and behavior of the tracks operating in the scenario, the ambiguity present in the stream of sensor reports, and the complexity of decisions faced by the correlation stage of the fusion system.

Most of the previously reported assessment approaches either focus on the performance issues related to individual tracks, such as track initiation probability and delay time[1], to individual algorithms[2], or attempt to characterize the improvement of operator performance when assisted by multi-sensor fusion versus individual sensor data streams[3]. Recently, interest has surfaced in the evaluation of relative performance of competing fusion solutions in the context of a fusion testbed[4]. Daum[5] reports an analytical method for bounding fusion performance in terms of the error covariance estimate. Boily[6] reports methods to evaluate tracking, identification, and global performance. His approach suggests a way of measuring track precision independent of ground truth. Our approach to performance evaluation addresses track picture quality as well as individual track fidelity. The approach reported in this paper supports performance validation, quantification, comparison, and prediction. Quantitative performance comparison also supports selection of fusion solutions and tuning of parameterized fusion systems.

ATL has successfully used the complexity and performance metrics to construct comprehensive sets of test cases, to evaluate test case complexity, and to measure the performance of competing fusion algorithms in the context of the Rotorcraft Pilot’s Associate Level 1 data fusion subsystem.

The Rotorcraft Pilot’s Associate (RPA) Advanced Technology Demonstration (ATD) is Army Aviation’s most ambitious science and technology program. Its objective is to apply artificial intelligence and state-of-the-art computing technologies to manage and integrate next generation mission equipment and battlefield information in order to enhance the lethality, survivability, and mission effectiveness of combat helicopters. The primary element of the RPA system is the Cognitive Decision Aiding Subsystem (CDAS), which performs situation assessment, planning, and cockpit information management. Since the potential utility of associate systems technology is wide ranging, the program is focused not only on individual helicopter platforms, but also on the requirements of warfighting commanders and the combined arms team. This Advanced Technology Demonstration program is managed by the Army Aviation Applied Technology Directorate. Boeing Helicopter Systems is the RPA prime contractor, Lockheed Martin Federal Systems is the major subcontractor, and Lockheed Martin Advanced Technology Laboratories is responsible for the real-time, compute-intensive Data Fusion Subsystem[7]. The data fusion system contains an innovative approach to the integration of classification data into the fusion process[8].

RPA real-time multi-sensor data fusion (DF) integrates inputs from large numbers of on-board and off-board sensors which describe ground and air targets as well as missiles. Mission scenarios are characterized by high target densities, high target maneuverability, rapid sensor update rates, and significant data uncertainties. Sensor errors and uncertainties affect kinematic and classification attributes received by DF. Sensors report track classifications with varying specificity.

Our experience has shown that the approach presented in this paper supports a thorough analysis of RPA fusion system performance covering the broad range of scenarios envisioned for the RPA reconnaissance and attack missions. The Data Fusion system has been integrated into the complete RPA system and is undergoing final tests in preparation for the RPA flight testing, which will begin later this year. Complexity analysis reduced the amount of testing required for performance validation and supported comparative evaluation of algorithms. In the future, we plan to develop mechanisms which dynamically adapt a fusion system to a changing environment with the help of the performance and complexity assessment techniques.

In Section 2 and 3 we describe the performance and complexity assessment methods in greater detail. Section 4 contains a description of the tests and the performance and complexity results obtained on the RPA fusion project. Section 5 concludes with our analysis and lessons learned of the performance assessment methodology.

2. PERFORMANCE ASSESSMENT

The performance of the Situation Assessment (Level 2), Threat Assessment (Level 3) and Process Refinement (Level 4) modules depend on the quality of the track picture created by the Level 1 multi-sensor multi-target fusion subsystem. Ultimately, enhanced operator performance, e.g. pilot performance in the RPA system, is the goal of the fusion suite. Performance assessment at Levels 2 and above are insufficient as a guide to system development, because they introduce extraneous variability associated with the display and control system, user training levels, etc. The performance metrics described in this paper can easily be related to operator performance and, at the same time, support thorough Level 1 fusion system evaluation.

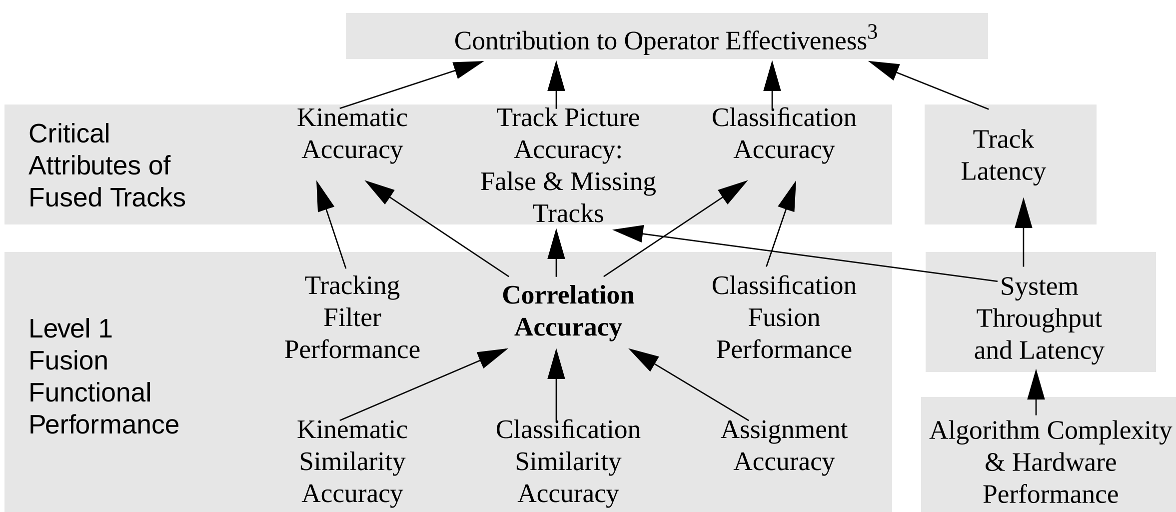

FIGURE 1. Correlation accuracy is the central contributor to Level 1 fusion performance.

Figure 1 shows the hierarchy of performance measures proposed. DF performance is measured by the accuracy and completeness of the fused tracks, which are output to the Situation Assessment module. The output tracks contain position and velocity (kinematic) estimates and classification hypotheses for each individual target track. A clean and accurate track picture generated by DF despite redundant and imprecise sensor reports is a prerequisite to superior pilot situational awareness.

2.1 Correlation Accuracy Metrics

Correlation accuracy is the central contributor to Level 1 fusion performance. In the correlation step, newly received sensor reports are correlated with fused tracks already held by the fusion system, i.e. the fusion system decides for each new report, which of the existing fused tracks it updates, or if it should initiate a new fused track. An error in this step significantly compromises the accuracy of subsequent processes, at least for a period of time following the error. Persistent track features, such as track classification and friend/foe indication, are especially vulnerable, because these attributes are propagated unchanged, unlike position and velocity. Major processing steps that precede correlation include track pre-processing, prediction, and clustering; major steps that follow correlation include fusion, i.e. the actual combination of the reported features with the existing track features - this includes track filtering, and post-precessing.

The correlation process consists of a set of similarity or gating functions and an assignment algorithm. Similarity functions measure how closely the newly received reports match the existing tracks to be extended; gating functions eliminate candidate assignments whose similarity falls below a threshold. The assignment algorithm determines a locally or globally optimal assignment of sensor reports to existing tracks. The RPA fusion solution employs the globally optimizing JVC (Jonker-Volgenant-Castanon) assignment algorithm. The development of fast but powerful, multi-dimensional similarity functions remains a research issue for the fusion community. Multi-dimensional similarity compares position, velocity, classification, identification, and other target features useful in distinguishing targets.

The correlation accuracy measure is the central metric of our performance assessment approach. It alone indicates most succinctly fusion system performance. Low correlation accuracy, i.e. correlation errors, inevitably corrupts all attributes of the fused track picture. Low performance of subsequent steps, such as state estimation and classification fusion, complicate the correlation problem and are likely to lead to correlation errors.

Correlation accuracy is measured at fusion system run-time by a small “built-in-test” code segment. It takes advantage of the availability of ground truth in the simulated scenarios. This measure is therefore unavailable during real operations. It is computationally cheaper to calculate and record the metric than to store all of the contributing data involved in their calculation. The correlation step may utilize all of the report and track attributes in an nxm comparison. The effort to extract and store all these features exceeds the effort to calculate a compact metric of correlation difficulty. Real-time metrics are essential enablers for the promising concept of run-time fusion system performance tuning.

2.2 Output Quality Metrics

The track picture and its constituent tracks are the externally observable outputs of the fusion system. They represent the attempts of the fusion system to reconstruct the ground truth scenario from sensor reports of varying quality and coverage. We have defined metrics which measure the quality of the instantaneous global track picture and the fidelity of the individual tracks which make up the track picture. Global track picture metrics evaluate the total number of tracks reported, the occurrence and persistence of false tracks, and the frequency with which tracks are missed. Individual track quality is measured by the distance of the reported target position to the actual position. The minimum, maximum, and average of the distances over the life of the tracks are computed. Track classification is evaluated by the accuracy and precision of the target classification. Classification precision measures whether the system correctly and effectively used target feature clues to narrow the set of possible platform classes.

Track picture metrics are sampled at regular time intervals and accumulated. The metrics evaluate the track picture as an instantaneous estimate of the true arrangement of targets at the sampling time. This approach does not attempt to evaluate the accuracy of kinematic track histories. Thus, if the fusion system were to mistakenly indicate that two tracks crossed, the error is only counted once when it is committed, even though its effects persist in the historical picture of the two tracks. On the other hand, errors in target classification or other persistent track attributes are detected and counted any time they are found in the fusion system output. Metrics on the fusion system output are calculated in non-real-time on the set of recorded fusion system output tracks.

2.3 Throughput Metrics

Fusion system throughput measures how many track updates can be processed in a given time period. Throughput impacts track picture reconstruction accuracy, because track updates must be skipped or output latency increases when system throughput is exceeded by the number of reports received. Track latency measures the delay until information about a track is handed off to the Situation Assessment module. Of most interest is the delay introduced by fusion subsystem processing. Fusion system throughput and latency depend on the performance of the computing hardware and on the complexity of the fusion algorithms.

Testing of the RPA Level 1 data fusion subsystem concentrated on measuring correlation accuracy, global track picture accuracy, and kinematic accuracy of the fused tracks on a multitude of test scenarios. Fusion subsystem throughput was also measured. Track latency is constant and determined by the fixed 10 Hz processing cycle of the RPA fusion solution.

3. COMPLEXITY ASSESSMENT

The accuracy of the track picture produced by Data Fusion depends on the quality of reports received from the sensor suite. Sensors performance, in turn, is influenced by the complexity of the ground truth scenario. All performance metrics must, therefore, be interpreted with respect to the level of complexity/difficulty posed by the test scenario. Figure 2 shows a dependency chart which relates errors in the fusion system output to their contributing factors.

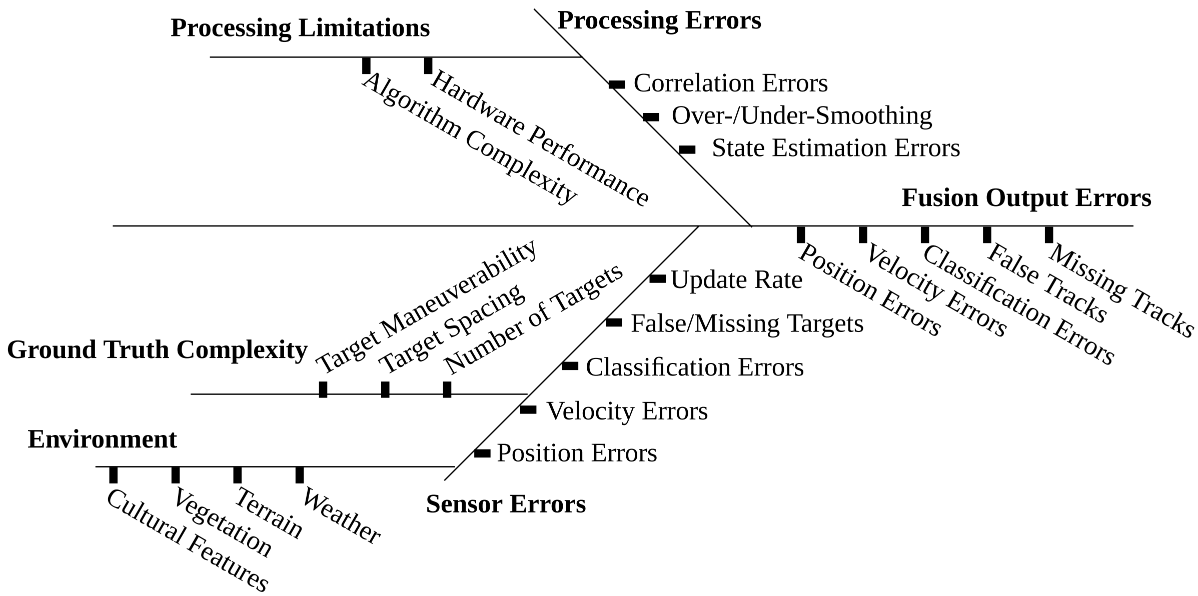

FIGURE 2. Fusion system output performance and errors can be explained from contributing factors.

ATL developed three complementary complexity metrics. One (Ground Truth Complexity) measures the test scenario directly, the second (Sensor Report Complexity), see Figure 3, measures the test scenario as seen through the stream of sensor reports, the third (Assignment Level of Difficulty) measures complexity of the correlation step during fusion module execution. Ground Truth Complexity predicts the complexity of correctly correlating track updates to fused tracks from the proximity and maneuverability of the ground truth targets. Sensor Complexity factors in sensor noise, i.e. errors in reported target position, and report intermittence. The Assignment Level of Difficulty is a DF internal measure which is calculated during DF execution and evaluates the actual difficulty of choosing the correct sensor update to CTF track assignments at each DF processing cycle. It focuses on the complexity faced by the assignment, i.e. the update-to-track association, component.

The approach assumes that sensor reports are compromised by errors whose statistics are known, and that reports are not intentionally misleading. For example, reported positions must approximate the distribution reported, for example, by means of an error ellipse. Scenario complexity changes over time. Therefore, the complexity measures generate an average of the instantaneous complexity for a given section of the scenario. Sections (intervals) are chosen to be long enough to smooth statistical variations.

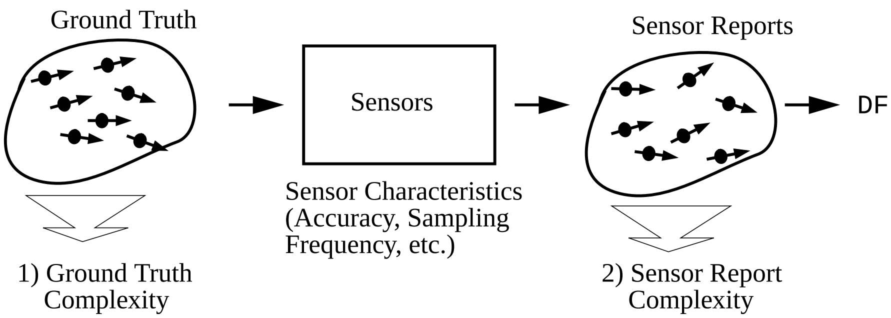

FIGURE 3. Ground Truth complexity is determined from target track attributes; Sensor Report complexity includes the effects of sensor limitations.

For RPA data fusion testing we selected five categories of test scenarios, see Section 4. Each of the five categories examines DF performance in the light of specific adversity. A “Braid” scenario contains multiple intersecting undulating tracks. A simulated “Mountain Pass” scenario stresses correlation algorithms with targets maneuvering in extreme proximity. A “Spiral” scenario simulates successively increasing report intermittence with multiple spiraling tracks. A “200 Track” scenario stresses DF throughput with 200 targets reported by five sensors with overlapping coverage. Scenarios within each test category vary individual target proximity, sensor positional accuracy, reporting intermittence, and/or target classification accuracy. Most scenarios were designed to be artificially and overly complex in order to collect enough errors to be able to draw valid conclusions.

3.1 Ground Truth Complexity

Ground Truth (GT) complexity is characterized by two factors: GT attribute complexity and GT time complexity. The effects of attribute complexity are felt due to sensor inaccuracies, and the effects of time complexity become critical because of sensor reporting intervals and intermittency. Calculation of both GT attribute and GT time complexity hinge on the definition of an appropriate report-to-report distance metric.

3.1.1 GT Attribute Complexity

The proposed measure of GT attribute complexity is based on the distribution of distances between attributes of GT targets, measured by a suitable distance metric and averaged over a chosen number of samples. It is not a single number but a representation which allows us to derive how many of the targets are closer to each other than a chosen threshold. This formulation expresses complexity relative to a particular resolution, i.e. the chosen distance threshold. This measure answers the question “*How difficult is it to distinguish the targets in the scenario*?”

Instantaneous GT attribute complexity is calculated from the instantaneous distance distribution by the formula

$$C_a(S_i, r) = \left(\sum_{1}^{n}(i-1)\right) \cdot \underset{i}{P}(\text{dist} < r) = \frac{n \cdot (n-1)}{2} \cdot \int_0^r p_i(\text{dist})d\text{dist}$$

where S_i is the GT scenario, r the chosen resolution, $p_i(\text{dist})$ is the distance distribution of the scenario, n is the number of targets in the scenario, and $n \cdot (n-1)/2$ is the number of target pairs.

We multiply the probability integral by the number of target pairs, in order to count the number of targets clustered within the resolution. This way, a scenario of 10 distances between 5 targets, where 5 distances are smaller than r, has complexity 5 instead of 0.5, and a scenario of 28 distances between 8 targets where 14 are close together has complexity 14 instead of 0.5 also. This formulation captures how many “difficult” decisions might have to be made selecting between close neighbors, if sensor reports directly represented ground truth targets. Figure 4 shows an example of a possible distance distribution.

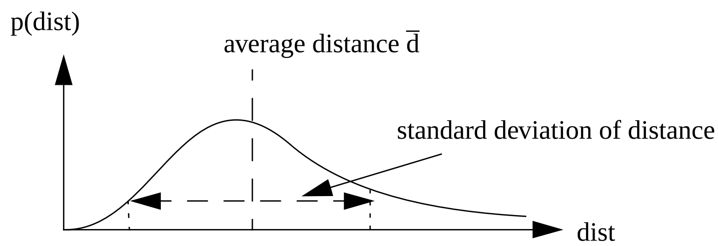

FIGURE 4. Example of a distance distribution p(dist) for a GT scenario.

For a specific GT scenario and resolution this quantity can be calculated as

$$C_a(S_i, r) = |\{(j, k)|(\mathrm{dist}_{j,k} < r), j, k = 1 \ldots n\}|$$

where $|\{(j, k)| \ldots\}|$ is the number of pairs of targets whose distance is smaller than *r*.

Figure 5 illustrates how the attribute complexity of the scenario depends on the chosen resolution. The attribute resolution *r* is a multidimensional quantity like the distance between (the attributes of) two targets. Later we will see how

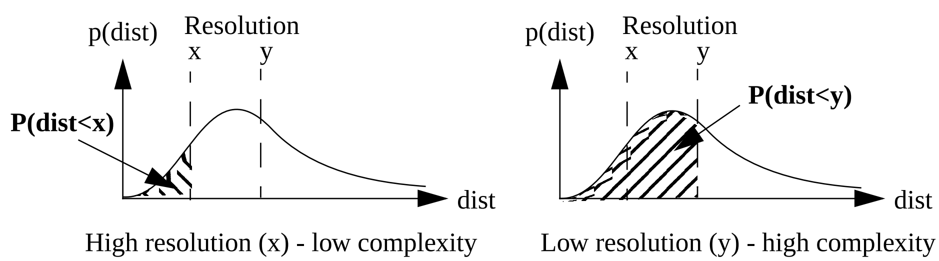

FIGURE 5. Complexity of a scenario for two different resolutions.

sensor characteristics determine resolution and thus allow to pin down GT complexity relative to sensor characteristics/resolution.

Interestingly, it is possible that, given two GT scenarios, one is less complex if the resolution is very high (small distances can be discerned) but becomes more complex for lower resolution, see Figure 5. Therefore, it is necessary to specify attribute resolution *r* before GT attribute complexity can be calculated.

3.1.2 Distance Metric

The complexity distance metric, like the similarity function used in the data fusion correlation process, has to measure how similar (or different) the attributes of two targets are. Target attributes include position, velocity, acceleration (commonly collected into a state vector), emission characteristics in the RF and optical frequency bands, exterior appearance, radar signature, etc. Typically, distances are measured independently for each of these dimensions and subsequently combined via a distance combination function. Any monotonic distance and distance combination function is acceptable for the proposed complexity measure. Problem specific methods must be developed for discontinuous attributes, such as radar signature. For RPA DF testing, we only used positional Euclidean distance to measure distance for the complexity metric.

3.1.3 GT Time Complexity

The proposed measure of GT time complexity is based on the distribution of the distances $p_i(\mathrm{dist}_\delta)$ between successive states of all targets over a nominal time increment δ, averaged over a suitable amount of time. The size of the

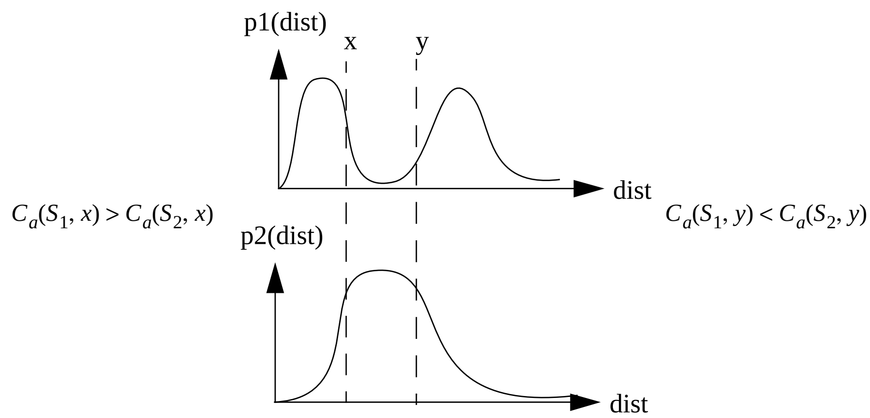

FIGURE 6. The relative complexity of two scenarios may depend on the resolution.

time increment is not critical as long as the changes that occur during the time intervals to be considered can be approximated by a linear extrapolation of the changes during the chosen increment. Of course, the time increments must be identical when the GT time complexity of different scenarios are to be compared. This measure answers the question "*How much does the scenario change over time*?"

GT time complexity captures a generalized (linearized) maneuverability measure of the targets, i.e. how rapidly target attributes change. High GT time complexity results from rapid change in target kinematic attributes, such as high speed, acceleration, and jerk (i.e. change in acceleration), and also changes in discrete attributes. For example, a target with multiple emitters which are operated independently can present a greatly different EW signature from one instant to another. Simply turning emitters on and off and changing emitter modes also contributes to generalized maneuverability.

The same distance function as is used to generate GT attribute complexity can be used here. GT time complexity must be expressed relative to time resolution, which is realized by sensor sampling intervals. Instantaneous GT time complexity is defined as the average of the instantaneous time-based distance distribution using the formula

$$C_t(S_i, \tau) = \int_0^{\infty} (p_i(\text{dist}_\delta) \cdot \text{dist}_\delta \cdot \tau)\, d\text{dist}_\delta = \tau \cdot \int_0^{\infty} (p_i(\text{dist}_\delta) \cdot \text{dist}_\delta)\, d\text{dist}_\delta = \tau \cdot \overline{\text{dist}}_\delta$$

where S_i is the GT scenario, τ the chosen time resolution, and $p_i(\text{dist}_\delta)$ the distribution of the incremental distances.

Therefore, it is necessary to specify the time resolution τ before GT time complexity can be calculated. However, unlike GT attribute complexity, GT time complexity is linear in τ and the complexity-ordering of scenarios does not depend on τ, i.e. if a scenario has higher GT time complexity than another at time resolution τ_1, it will also have higher complexity at any other arbitrary time resolution τ_2.

3.1.4 Total GT Complexity

Total GT complexity is a function of GT attribute and GT time complexities.

A possible (rough) measure of total GT complexity is

$$C(S_i, \tau) = C_a(S_i, (r = C_t(S_i, \tau))) = C_a(S_i, (\tau \cdot \overline{\text{dist}}_\delta))$$

where attribute complexity is evaluated at a resolution determined by time complexity. This measure answers the question "*How relevant are the changes in the scenario over a specified amount of time, i.e. are the changes in target attributes large enough to impact target distances*?"

This measure is a compromise which assumes that target attributes are evenly distributed. If, for example, a scenario consists of one set of slowly changing targets that are far apart and another set of quickly changing targets that are close together, the proposed measure of total GT complexity will underestimate the complexity of this scenario. A more precise measure has to combine distance and maneuverability for each target separately and aggregate the results for all targets in the scenario. This is analogous to the current Assignment Level of Difficulty metric.

A more precise measure of total GT complexity for a specific resolution is

$$C(S_i, \tau) = \left|\{(j, k) | (\mathrm{dist}_{j,k} < \tau \cdot (\mathrm{dist}_{\delta, j} + \mathrm{dist}_{\delta, k})), j, k = 1 \ldots n\}\right|$$

where $|\{(j, k)|\ldots\}|$ is the number of inter-target distances which can be exceeded by the averaged maneuverabilities $\mathrm{dist}_{\delta, j}$ and $\mathrm{dist}_{\delta, k}$ of targets j and k over time τ.

3.2 Sensor Report Complexity

The proposed metric of sensor report (SR) complexity measures how difficult it is to construct continuous tracks from often sporadic, discontinuous sensor reports. It is based on the concept of the *report-to-report variation* between successive reports. It applies equally to single and multi-sensor scenarios, continuous and intermittent sensor reports, and varying amounts of attribute information provided by sensor reports. It can be measured knowing just the GT identity for each report and without knowing GT attributes.

3.2.1 Limitations

The proposed SR complexity measure does not measure how far the sensor reports deviate from GT but only how separable, consistent, and continuous the reports are with respect to each other. If a sensor reports position with a constant offset, however large, and no noise, its consistency is very high and its variation low; thus, the resulting SR complexity will be low, even though the GT cannot be reconstructed precisely from the sensor reports due to the positional offset. An additional, separate measure of attribute, e.g. positional, bias should capture systematic GT to sensor report differences, i.e. SR bias.

3.2.2 Report-To-Report Variation

Report-to-report variation is the change in attribute values between successive reports on the same GT target, regardless of which sensor supplied the report. Report-to-report variation is analogous to GT time complexity, where the time resolution τ is now determined by the interval between successive sensor reports, and sensor errors add to apparent target maneuverability.

Variation between successive reports from one sensor or between coincident reports from multiple sensors create opportunity for error. The ideal sensor would report infinitely fast and perfectly accurately, so that variations between successive reports become infinitesimally small. Variations are caused by sensor imperfections and by target maneuverability, which is the more detrimental the lower the reporting rate. For the purpose of measuring SR complexity, it is irrelevant what caused the variation, be it target maneuverability, unequal bias among multiple sensors, or sensor noise.

Instantaneous report-to-report variation is defined as the change over time in reported attribute values $\mathrm{dist}_j(T_{k,j}, T_{k-1,j}) / (T_{k,j} - T_{k-1,j})$ between the two most recent reports on target *j* received at times $T_{k,j}, T_{k-1,j}$. *Target report variation,* $trv_j(T_{k,j})$, is the average of instantaneous report-to-report variation calculated over a sliding time window for each target.

Missing Attribute Values. Missing (unreported) attributes require special processing. We propose to maintain a moving average of the differences between values of specific attributes from report to report, and to substitute this average for the actual difference when attribute values are not reported in either or both of the reports which are being compared. In the average calculation, a difference involving a missing attribute is considered to be zero. Initially, the average difference is set to zero. It remains zero if the attribute is never reported.

3.2.3 Correlation Complexity

Correlation complexity quantifies the difficulty of correctly correlating sensor reports with existing fused tracks despite target maneuverability, sensor noise, and sensor intermittence. The complexity calculation does not maintain fused tracks but estimates the difficulty directly from the stream of sensor reports. It predicts for different scenarios the relative number of assignment errors committed by a fusion algorithm operating on the stream of reports. Actual assignment errors depend on the performance of the fusion algorithms. The difficulties associated with establishing new tracks and dropping terminated tracks are addressed only indirectly by their effects on the assignment process.

The proposed measure assumes that sensors deliver at time T_k new reports $R_j(T_k)$ to the fusion system in batches of *m* reports. The *m* reports may include false target reports. The measure assumes to also have access to the latest report $R_j(T_{k-1,j})$ on each of the true targets $j = 1 \dots n$, regardless of which sensor reported it, but does not presume the existence of a fused target track. Time $T_{k-1,j}$ is not a fixed instant in time but the last time a report for target *j* has been received.

The calculation is processed in four steps. In Step One, a confusion set $F_j(k)$ is generated for each target *j* updated in $R_j(T_k)$. This set contains all the reports from the set $R_j(T_k)$ which are within the target report-to-report variation of the updated target *j*.

$$F_j(k) = \{ i | (\text{dist}(R_j(T_{k-1,j}), R_i(T_k)) < tr\text{v}_j(T_{k,j}) \cdot (T_{k,j} - T_{k-1,j}), i = 1 \dots m) \}$$

In Step Two, we determine the number of confusion sets that each report *i* falls into.

$$NF_i(k) = \left| \{ j | (i \in F_j(k), j = 1 \dots m) \} \right|$$

In Step Three, we calculate the correlation complexity for one update cycle, i.e. for one batch of sensor reports.

$$C_{c,\,\text{sensor}}(k) = \sum_j \left(\frac{\sum_{i \neq j,\, i \in F_j(k)} 1/NF_i(k)}{\sum_{i \in F_j(k)} 1/NF_i(k)} \right)$$

In Step Four we calculate scenario correlation complexity by summing over the correlation complexity per update for the length of the scenario. The scenario correlation complexity calculates the number of chances for correlation errors.

$$C_c = \sum_k C_{c,\,\text{sensor}}(k)$$

Scenario correlation complexity serves as the Sensor Report (SR) complexity metric.

3.3 Assignment Level of Difficulty

Assignment Level of Difficulty (ALoD) is calculated as the data fusion system is executing. ALoD is measured for each fused track when similarity (or cost) functions are computed for that fused track and the new sensor reports. For a given fused track, if a sensor report exists with the same target number as that fused track, then the difficulty is computed using the algorithm shown below. Otherwise, the difficulty for the fused track is defined to be the number of sensor reports.

When an assignment is made of a sensor report to a fused track, the difficulty for that CTF track is added to a cumulative total of difficulty for all assignments for that fused track, as well as to a cumulative total of the difficulty of all assignments made by Data Fusion. If the assignment was an error, the difficulty is also added to cumulative totals of difficulty for erroneous assignments for both the fused track and all of Data Fusion.

Assignment Level of Difficulty calculates a measure of confusion possibilities for a particular fusion system, which is analogous to the generic metric of error possibilities that constitutes Sensor Report complexity. ALoD predicts assignment errors committed by the fusion algorithm more precisely than the SR complexity metric which is based only on the stream of sensor reports.

$$\Phi = \text{Set of sensor reports}$$

$$\text{Cost}_{\text{S,C}} = \text{Cost of assignment of sensor report S to fused track C}$$

$$\text{Correct}_C = \text{Cost of assignment of correct sensor report to C}$$

$$P_T = \text{Player number of sensor report or fused track T}$$

$$Z_{\text{S,C}} = \exp\left(K \times \left(\frac{(\text{Correct}_C - \text{Cost}_{\text{S,C}})}{\text{Correct}_C}\right)\right)$$

$$\text{Difficulty}_C = \sum_{S \in \Phi,\, P_S \neq P_C} Z_{\text{S,C}}$$

4. TEST SCENARIOS AND RESULTS

The performance and complexity assessment techniques described above supported development, tuning, and validation of the RPA data fusion subsystem. Tests were executed using an instrumented version of the DF code and using test data from an input simulator which attaches ground truth tags to the sensor reports. Knowledge of ground truth is necessary to judge the correctness of the correlation and assignment process. Each scenario was executed multiple times and the resulting performance data were averaged. A suite of analysis tools was used to calculate aggregate performance measures from the data collected. Ground truth (GT) and Sensor Report (SR) Complexity were determined before each run. Assignment level of difficulty, assignment errors, and the Central Track File (CTF) output, i.e. fused track oputput, were captured during each run. Assignment errors determine correlation performance. The differences between ground truth tracks and CTF tracks determine kinematic and classification accuracy. In cases where multiple CTF tracks approximate a GT track or where updates for one GT track were alternately associated with two or more CTF tracks, the CTF track which approximated the GT track most closely was used for comparison. Other CTF tracks are false tracks or represent a different GT track.

Figure 7 and Figure 8 summarize the major results of DF performance testing. Figure 7 illustrates the sensitivity of DF to two of the most critical scenario parameters: GT inter-target distance and sensor positional error. The table shows that more closely spaced targets need to be observed with more precise sensors in order to get optimal perfor-

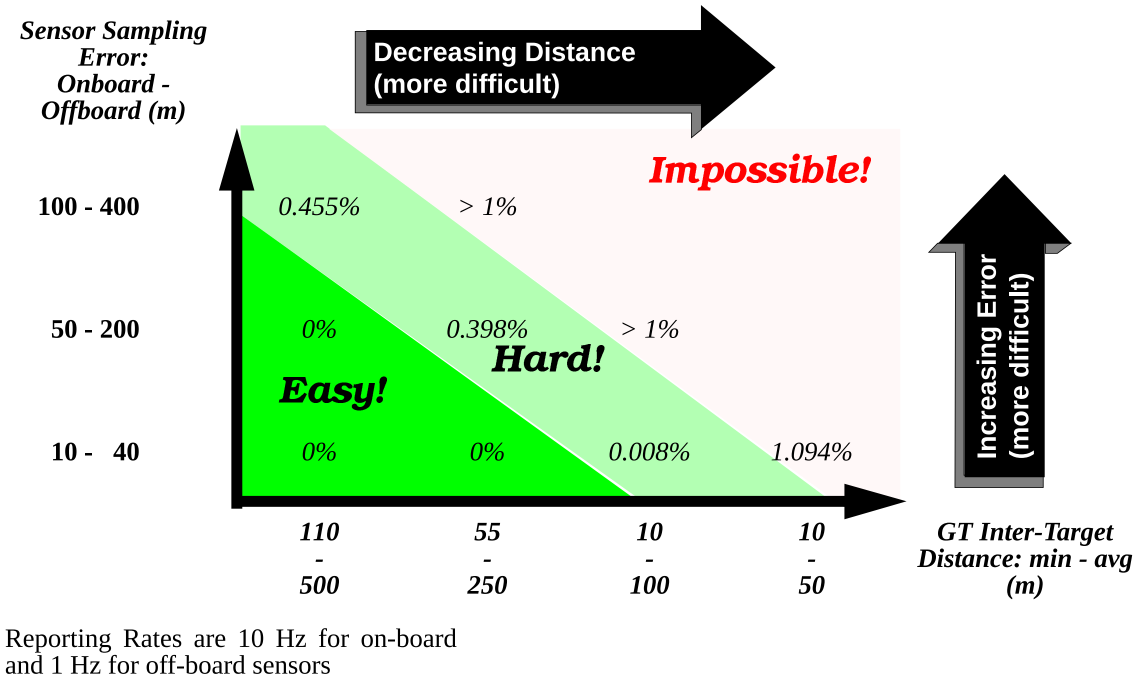

FIGURE 7. Systematic performance assessment reveals thresholds of data fusion system performance.

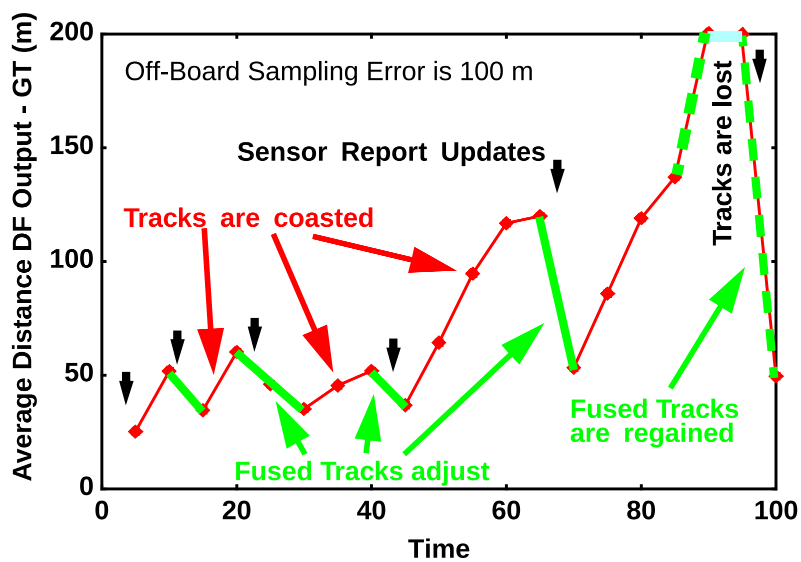

FIGURE 8. Systematic performance assessment illustrates critical data fusion system behavior and performance limitations.

mance. Nevertheless, DF tolerates occasional errors that are larger than the minimum target separation. Similar results were obtained with varying report intermittence.

Figure 8 illustrates the kinematic accuracy of a set of six tracks when reports are becoming more intermittent as indicated by the dotted lines. Every report received realigns the CTF tracks with the ground truth. During intervals when no reports are received, the tracks are coasted in straight lines and drift away from the GT tracks. GT tracks travel along curved paths with constant radii of 50, 100, 200, 500, 1000m, and infinity at a speed of 10 m/s.

In summary, system requirements were proven to be satisfied. The test methodology described here has been shown to correctly predict DF performance within the context of the RPA system.

4.1 Methodology for Generating Scenarios

In order to adequately test any Data Fusion system, a wide range of scenarios is required so that performance can be evaluated under all of the conditions which may impact the performance of the fusion system. Our methodology for testing the Data Fusion system developed under the Rotorcraft Pilot's Associate program included the following steps:

1. Identification of all conditions or combinations of conditions which may impact the performance of the fusion system. These are described in section 5.2 below.
2. Definition of a prototype scenario for each condition or combination of conditions to be tested. A prototype scenario includes the number and type of target entities, the approximate trajectories, the length of the scenario, and the number and type of sensors providing data to the type data fusion system. Use the Data Fusion Input Simulator (DFIS), described below, to generate the basic scenario.
3. For each prototype scenario, select one or more scenario parameters to vary to test the conditions for which the scenario was selected, and the range for the selected parameters. For example, if the condition to be evaluated were the homogeneity or heterogeneity of class information in the incoming sensor data, the scenario parameters to be varied might include whether or not the sensors could determine and report class information about the targets.
4. For each combination of values for the selected parameters, use the Data Fusion Input Simulator to generate data files for input to Data Fusion representing the sensor input to data fusion from the scenario with that combination of parameters.
5. For each variation of a single prototype scenario, run Data Fusion with the input created for that variation, and collect the data described in sections 2 and 3.

Test emphasis was placed on further quantifying the range of scenarios and sensor configurations within which Data Fusion performs reliably and accurately. Scenario parameters of interest include the number of targets, the separation of targets, and target maneuverability. Sensor parameters of interest include the mix of active sensors, sensor accuracies, sensor intermittence, and sensor reporting rates. Testing identified how close DF is to meeting the performance requirement of processing 200 Central Trackfile tracks at realistic maximum sensor input rates.

4.1.1 Data Fusion Input Simulator

In the development of a sensor Data Fusion solution for RPA, it is necessary to stimulate that subsystem with both realistic and overly stressful sensor scenarios. The Data Fusion Input Simulator (DFIS) is a user-friendly engineering tool designed for expedient creation of these on-board and off-board scenarios. The DFIS provides a means for RPA Data Fusion subsystem development and stand-alone subsystem validation and verification. The DFIS was designed with the intention of permitting an engineer to quickly and easily generate battlefield scenarios consisting of air and ground vehicles. These scenarios may be developed in two ways: with a graphical drawing window that allows the user to view a scenario as it unfolds, or non-graphically, by manipulating data files.

In either mode, the user creates a scenario by specifying overall scenario characteristics, such as the duration of the scenario and the maximum number of battlefield entities or players. The trajectory of each player is defined as a series of waypoints, and the vehicle type is selected from a pre-defined taxonomy. In addition to player entities, the user defines the set of sensors which can provide data to the data fusion system. There are a total of 21 possible sensors or other data sources available to provide input to the data fusion system in RPA, including an onboard Target Acquisition System (TAS) - a MMW radar and FLIR combination, an onboard passive RF sensor, an onboard Laser Warning Receiver, and a number of offboard sources including JSTARS, AWACS, and TAS and RF sensor data from up to three wingman aircraft. Each sensor data source can be positioned independently or made to move with one of the defined players, and set to operate in a desired mode, which can include Tracked Reports, Untracked Reports, Bearing Only Reports, and Group Tracks, depending on the sensor. Other parameters, such as update rate, positional error, probability of detection per target type, and classification capability, can be set for each sensor. The result is a very powerful capability to specify all of the characteristics of the data that will be made available to Data Fusion.

DFIS generates two levels of scenario data. The first, Ground Truth, includes the true position, velocity, and other properties of each player at 10 Hz intervals for the duration of the scenario. The second, sensor data, is generated from the ground truth and consists of the sensor reports generated by the defined sensor data sources, whose timing and characteristics depend on the parameters set for each sensor and on the underlying ground truth data. The ground truth and sensor data are stored in a binary output file which is read by Data Fusion and by other software developed to support performance analysis.

4.2 Types of Scenarios Generated

TABLE 1. Desired Conditions for Performance Analysis and Expected Impact On Data Fusion Performance

CONDITIONS TO VARY FOR ANALYSIS	EXPECTED IMPACTS ON DATA FUSION PERFORMANCE
• Target Separation • Sensor Errors	Increase in Data Fusion Error Rate with decreasing target separation or increasing sensor error.
• Sensor Distance • Presence or Absence of Class Data	Increase in Data Fusion Error Rate with increasing sensor range to target or absence of class data.
• Target Maneuverability • Sensor Data Intermittency	Increase in Data Fusion Error Rate with increasing target maneuverability or sensor data intermittency.
• High Target Volume	Increasing error as Fusion is unable to keep up with data volume.

Table 1 describes the sets of conditions for performance analysis and expected impact on the performance of the Data Fusion system.

Five types of test cases were created and evaluated. The five types of test cases are:

1. “Braid”: Group 1 of test cases (cases 1.1 through 1.17) are based on four variants of a scenario of ten vehicles moving in close proximity on separate but intersecting sinusoidal tracks. Average target separation was set to four different values: 50 m, 100 m, 250 and 500 m. Sustained minimum inter-target distances were 11, 10, 55, and 110 m. TAS and/or JSTARS report at varying rates. Sensor errors were varied. The effects of target class data and of sensor tracking and filtering were also studied using these scenarios.
2. “Mountain Pass”: Group 2 of test cases (cases 2.1 through 2.6) model a group of 35 targets moving through a mountain pass in dense formation with TAS and JSTARS reporting. Three different sensor configurations are exercised: JSTARS at 25 km and TAS at 2 km, JSTARS at 100 km and TAS at 2 km, and JSTARS at 250 km and TAS at 5 km. Results of DF performance with and without using class in the track-to-track association process, i.e. in the cost function, are presented.
3. “Spiral”: Group 3 of test cases (case 3.1 and 3.2) analyzes the effects of target maneuverability and sensor report intermittency with bursts of reports separated by increasing intervals where no reports are received.
4. “Mission”: Case 4 is inspired by scenarios used for the official RPA evaluations. TAS (with IFF), Team Member TAS, EOB, JSTARS, and ASE sensors are active at different times.
5. “200 Track”: Case 5 determines maximum throughput performance of DF with a scenario of 200 targets moving toward ownship in five groups of 40 vehicles each. A JSTARS and a TRIXS sensor are active. Each reports 60 targets every second, which is close to the maximum input rates for these two data sources. The onboard TAS and AEOCM sensors report 80 and 20 targets, respectively, at a 10 Hz rate. Each sensor scans the set of targets for a period of three minutes.

5. CONCLUSIONS

Our systematic approach to data fusion system performance testing has proven to be an invaluable tool for system development, tuning, and validation. We have established limits on the applicability of the fusion algorithms developed for RPA and have shown that the expected RPA mission scenarios fall within these performance limits. For example, as shown in Figure 7, the RPA data fusion solution performs nearly error-free given onboard and offboard sensor errors of 10 and 40 m, respectively, as long as the inter-target distance stays above a minimum/average of 10/100 m. The chart displays this value and the corresponding zero-error boundary, which can be drawn from test results on scenarios with appropriately varied parameters. The one-percent error boundary of the RPA fusion system is shown. too. Additional performance charts not shown here reveal constant error boundaries of fusion performance relative to report update frequency and sensor error as well as to the combination of update frequency and inter-target distance. These charts are the results of numerous, repeated tests on a large set of test scenarios.

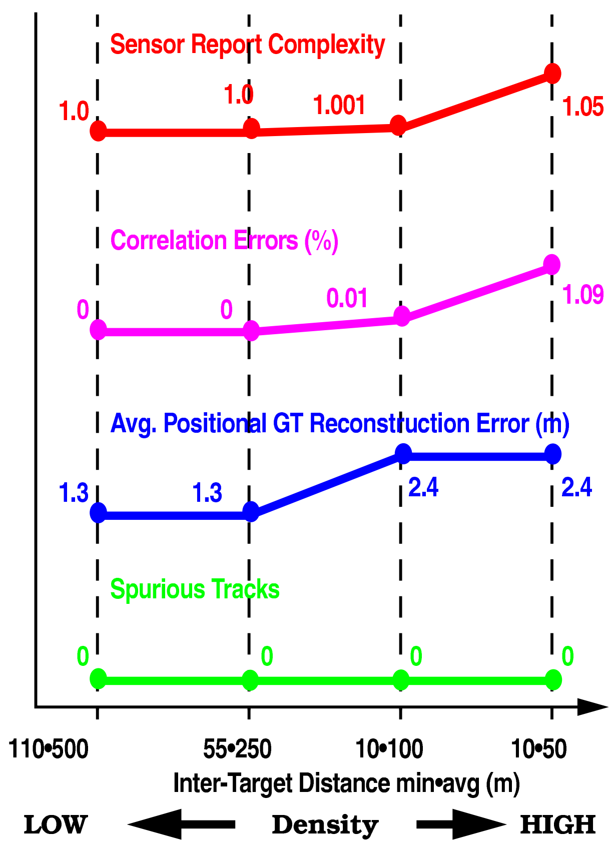

FIGURE 9. Ground Truth reconstruction performance and correlation errors are proportional to Sensor Report complexity. (Results from the Mountain Pass scenario)

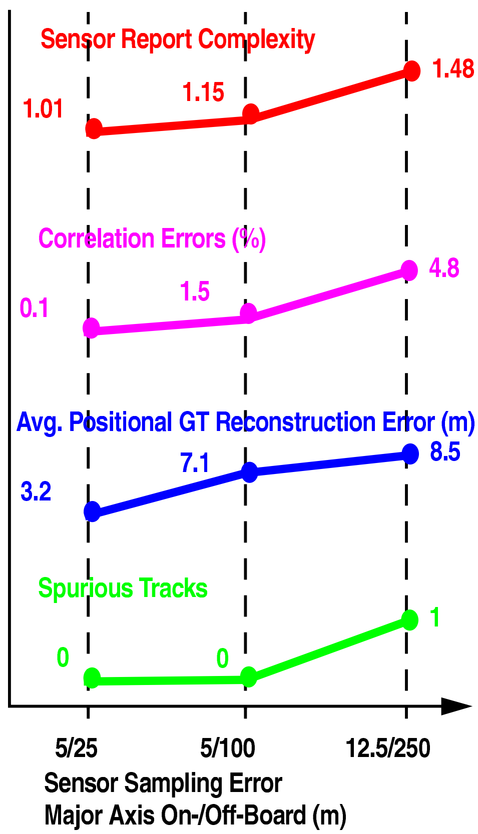

FIGURE 10. Results from the Braid scenario corroborate the predictive power of Sensor Report complexity.

With the approach described in this paper we can determine fusion system applicability without the cost of executing the fusion system on a multitude of scenarios. Fusion system applicability can be determined directly from an analysis of the characteristics of the expected scenarios via the Ground Truth and Sensor Report Complexity measures presented above. We have constructed reliable metrics which anticipate the sensitivity of the fusion system to the fundamental scenario characteristics, such as inter-target distances, sensor errors, and sensor reporting rates. Figure 9 and Figure 10 below show that the Sensor Report Complexity metric accurately predicts assignment errors and subsequently ground truth reconstruction performance. The results presented are based on measurements on the "Mountain Pass" and the "Braid" scenario, respectively. As shown in Figure 9, the point at 10/100 m minimum/average inter-target distance, where correlation errors first appear, is predicted exactly by the sensor report complexity measure. In general, it can be observed that the correlation error curve follows the Sensor Report Complexity curve faithfully.

Ground truth reconstruction error does not continue to increase with Sensor Report Complexity and correlation errors for the following reason. Complexity and correlation errors increase with decreasing inter-target distance, as shown in Figure 9, because the sensor updates for multiple fused tracks become kinematically indistinguishable from each other, i.e. they all fall within a tight neighborhood of the actual target position. Therefore, the wrong sensor report still represents a good approximation of the actual target position and the kinematic quality of the fused track remains unchanged despite the correlation error. On the other hand, when correlation errors become more numerous due to increased sensor error, as in Figure 10, the ground truth reconstruction error keeps increasing, simply because the sensor reports fall farther from the actual ground truth target position.

In the future we plan to implement classification accuracy and precision metrics and to re-target the performance assessment methodology towards run-time fusion system tuning.

REFERENCES

1. Cassassolles, E. and Tomasini, B., "Performance Evaluation of Multiple Hypothesis Filter for Track Initiation," Proc. SPIE Vol. 2235, pp. 441-447, , Signal and Data Processing of Small Targets 1994, Oliver E. Drummond; Ed.
2. Malkoff, D., "Evaluation of the Jonker-Volgenant-Castanon (JVC) Assignment Algorithm for Track Association", Proc. SPIE Vol. 3068, pp. 228-239, Signal Processing, Sensor Fusion, and Target Recognition VI, Ivan Kadar; Ed.
3. Stiles, P. and Hofmann, M.O., "Demonstrated Value of Data Fusion and Situation Assessment," American Helicopter Society Avionics Specialists Meeting, Sept. 23-25, 1997, Philadelphia, PA.
4. Hlavaty, C. and Waltz, E., "Integration Standards and Performance Metrics for Next Generation Avionics Fusion Architecture," Proc. IRIS Sensor & Data Fusion, 1997, Volume 1, pp. 199-208.
5. Daum, F.E., "Cramer-Rao bound for multiple-target tracking," Proc. SPIE Vol. 1481, pp. 341-344, Signal and Data Processing of Small Targets 1991, Oliver E. Drummond; Ed.
6. Boily, E., "Performance Evaluation of a Multi-Sensor Data Fusion Implementation - A Progress Report," Proc. SPIE Vol. 2235, pp. 570-579, Signal and Data Processing of Small Targets 1994, Oliver E. Drummond; Ed.
7. Malkoff, D. and Pawlowski, A., "RPA Data Fusion", 9th National Symposium on Sensor Fusion, Vol. 1, Infrared Information Analysis Center, pp 23-36, September 1996.
8. Hofmann, M.O., "Multi-Sensor Track Classification in Rotorcraft Pilot's Associate Data Fusion", Proc. American Helicopter Society Forum 53, Virginia Beach, VA, April 1997.